Mon. Not. R. Astron. Soc. **380,** 1608–1620 (2007) doi:10.1111/j.1365-2966.2007.12203.x

Empirical photometric redshifts of luminous red galaxies and clusters in the Sloan Digital Sky Survey

P. A. A. Lopes[1,2]★

[1]*Instituto Nacional de Pesquisas Espaciais – Divisão de Astrofísica (CEA), Avenida dos Astronautas, 1758 São José dos Campos, SP 12227-010, Brazil*
[2]*Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*

Accepted 2007 July 6. Received 2007 July 2; in original form 2007 February 8

ABSTRACT

In this work I discuss the necessary steps for deriving photometric redshifts for luminous red galaxies (LRGs) and galaxy clusters through simple empirical methods. The data used are from the Sloan Digital Sky Survey (SDSS). I show that with three bands only (*gri*) it is possible to achieve results as accurate as the ones obtained by other techniques, generally based on more filters. In particular, the use of the $(g - i)$ colour helps improving the final redshifts (especially for clusters), as this colour monotonically increases up to $z \sim 0.8$. For the LRGs I generate a catalogue of ∼1.5 million objects at $z < 0.70$. The accuracy of this catalogue is $\sigma = 0.027$ for $z \leqslant 0.55$ and $\sigma = 0.049$ for $0.55 < z \leqslant 0.70$. The photometric redshift technique employed for clusters is independent of a cluster selection algorithm. Thus, it can be applied to systems selected by any method or wavelength, as long as the proper optical photometry is available. When comparing the redshift listed in literature to the photometric estimate, the accuracy achieved for clusters is $\sigma = 0.024$ for $z \leqslant 0.30$ and $\sigma = 0.037$ for $030 < z \leqslant 0.55$. However, when considering the spectroscopic redshift as the mean value of SDSS galaxies on each cluster region, the accuracy is at the same level as found by other authors: $\sigma = 0.011$ for $z \leqslant 0.30$ and $\sigma = 0.016$ for $030 < z \leqslant 0.55$. The photometric redshift relation derived here is applied to thousands of cluster candidates selected elsewhere. I have also used galaxy photometric redshifts available in SDSS to identify groups in redshift space and then compare the redshift peak of the nearest group to each cluster redshift. This procedure provides an alternative approach for cluster selection, especially at high redshifts, as the cluster red sequence may be poorly defined.

Key words: surveys – galaxies: clusters: general – galaxies: distances and redshifts.

1 INTRODUCTION

Recent galaxy redshift surveys [2dF Galaxy Redshift Survey (2dFGRS), Colless et al. 2001; Sloan Digital Sky Survey (SDSS), York et al. 2000] have provided the astronomical community a unique view of the local universe ($z \sim 0.1$). Such surveys are based on spectrographs that simultaneously observe hundreds of objects. Although the improvement respective to a decade ago is enormous, larger and mainly deeper general spectroscopic surveys are not yet possible with current instrumentation. Note that surveys targeting specific populations, such as luminous red galaxies (LRGs) (SDSS, Eisenstein et al. 2001) or star-forming galaxies (Wiggle-*z*, Glazebrook et al. 2007), can sample much larger volumes. For the mean time, photometric redshifts provide a valuable alternative to probe faint sources within large areas.

★E-mail: paal05@gmail.com

Photometric redshift techniques are essentially a mechanism to convert photometric properties of galaxies (such as colours) into redshift and physical properties (e.g. luminosity and type). Thus, with the proper choice of passbands and the use of an accurate photometric redshift algorithm, it is possible to map the distant universe in three dimensions. These surveys represent a powerful tool for studying the statistical properties of galaxies and their evolution.

There are several photometric redshift estimators developed to date. These can be generally classified either as empirical or template-based methods. In the first case a direct relation is obtained through the comparison of the photometric properties (colours) and spectroscopic redshifts. Such empirical relations can be derived, for instance, through polynomial fitting (Connolly et al. 1995) or neural networks (Collister & Lahav 2004). The template-based algorithms rely on the availability of a set of galaxy templates. These should accurately represent the distribution of galaxy spectral energy distributions (SEDs) and their evolution with look-back time (Csabai et al. 2003). Hybrid photometric redshift techniques have

also been proposed in the last few years. They combine the advantages of empirical and template-fitting methods by iteratively improving the concordance between photometric data and the SEDs. In other words, the template spectra are reconstructed to best match the observed photometric measurements of each galaxy (Budavári et al. 2000; Csabai et al. 2003).

LRGs are marked by uniform SEDs, characterized by a strong break at 4000 Å due to the accumulation of a number of metal lines. The shift of this feature through different filters is strongly correlated with redshift. These galaxies are also known to be some of the most luminous objects in the universe and are preferentially found at high-density environments, rendering these objects an interesting tool for selecting and studying clusters. All that said, it is clear that LRGs comprise an optimal population to derive accurate photometric redshifts to very large distances.

This paper describes the construction of a large photometric redshift catalogue of LRGs at $z < 0.70$. This catalogue is based on simple polynomial fitting of the relations between galaxy colours and spectroscopic redshifts. I explore the use of different colours from SDSS, showing that with three bands only it is possible to achieve results comparable to more elaborated empirical techniques, such as ANNz (Collister & Lahav 2004), kd-trees or the nearest neighbour method (Csabai et al. 2003). In addition, I employ similar relations to derive photometric redshifts of galaxy clusters.

When estimating redshifts of clusters the main drawback is the need to apply a background correction when selecting probable cluster galaxies seen in two dimensions. I discuss different possibilities when minimizing the background effects, showing that the most precise results can be achieved when selecting the reddest galaxies (the selection is based on the $u - r$ colour). Photometric redshift estimates of clusters rely on precise values of their median colour. Elliptical or S0 galaxies comprise the main population in the central regions of galaxy clusters. Thus, one would like to use these galaxy types when estimating the typical colours of clusters. As it is shown in Section 4.2, at low redshifts a simple statistical background correction is enough to minimize the influence of galaxies that do not belong to the clusters, and to accurately estimate cluster colours. However, at higher redshifts this simple correction leads to an underestimation of cluster colours compared to the expected values for ellipticals. That is also due to the increase with redshift in the fraction of blue galaxies in clusters (Butcher & Oemler 1984). To circumvent this problem, the use of the $u - r$ colour plays a key role to help the selection of early-type systems and thus reduce the scatter of the observed colours of clusters. More details are found in Section 4.2.

This paper is divided as follows. In the next section I describe the SDSS survey, which is used as the basis for obtaining the empirical relations and evaluate the results. In Section 3 I describe the selection of LRGs and the photometric redshift technique employed for these objects. The same is done for clusters in Section 4, where I also make considerations about redshift accuracy. I also use galaxy photometric redshifts from SDSS for the identification of groups in redshift space. I summarize the results in Section 5. Throughout this paper I assumed a cosmology with $\Omega_{\rm m} = 0.3, \Omega_\Lambda = 0.7$ and $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, with h set to 0.7.

2 DATA

The photometric and spectroscopic data for this paper were taken from the fifth release of the SDSS (York et al. 2000). The SDSS consists of an imaging survey of π sr of the northern sky in five optical passbands (*ugriz*), from 3500–8900 Å. This will provide photometry for of the order of 5×10^7 galaxies. Spectroscopic survey will provide redshifts and spectra for $\sim 10^6$ of these. The survey is carried out using a 2.5-m telescope, an imaging mosaic camera with 30 CCDs, two fibre-fed spectrographs and a 0.5-m telescope for the photometric calibration. The imaging survey is taken in drift-scan mode and the data are processed with a photometric pipeline (PHOTO) specially written for the SDSS data.

Targets for spectroscopy are selected by the targeting pipeline from the imaging. Spectroscopic fibres are assigned to the targets by a tiling algorithm (Blanton et al. 2003). The minimum distance of 55 arcsec between the fibres leads to a loss of $\sim$6 per cent of galaxies, which is the main source of incompleteness.

The spectroscopic survey is originally divided in three samples. The 'main', flux-limited sample, has a median redshift of 0.104 and a limiting magnitude of $r_{\rm petro} \sim 17.77$ (Strauss et al. 2002). As this limit is much brighter than that for the imaging, the redshift completeness is nearly 100 per cent. The second sample is the LRG sample, which is approximately volume limited to $z \approx 0.38$ (Eisenstein et al. 2001), extending to $z \approx 0.55$. Finally, the quasar sample is defined by objects with colours distinct from those of ordinary stars. The completeness of this sample depends somewhat on redshift. In particular, the completeness is low for $2.4 < z < 2.9$, where the quasar and stellar loci cross; it is similarly low at redshifts around 3.5 and 4.5.

In addition to these data, the 2dF-SDSS LRG and Quasar Survey (2SLAQ; Cannon et al. 2006) has been recently completed. This survey exploits the high-quality SDSS imaging combined with the extraordinary spectroscopic capabilities of the 2dF instrument on the 3.9-m Anglo-Australian Telescope. It then results in a spectroscopic redshift catalogue of $\sim$13 000 LRGs at $0.4 < z < 0.7$.

These data can be found in the data release five (DR5) of SDSS or directly from the 2SLAQ web site.[1] In this paper I use photometric data from SDSS and spectroscopic redshifts from SDSS and 2SLAQ surveys (available within SDSS).

All the data selected from SDSS are from the DR5. I have selected only objects from the 'Galaxy' view (so that only PRIMARY objects are allowed) in order to avoid duplicate observations. Standard flags for clean photometry are also enforced. When selecting spectra and imaging, a *joined* query of the Galaxy and SpecObj (objects with clean spectra) views is performed. All the magnitudes retrieved from SDSS are dereddened (corrected for extinction) model magnitudes.

3 SELECTION OF LUMINOUS RED GALAXIES

The selection criteria adopted for constructing a photometric sample of LRGs is analogous to the description given in Padmanabhan et al. (2005), which is aimed at selecting a uniform sample of LRGs at $0.2 < z < 0.7$. Two different criteria are applied for selecting a low-redshift sample (Cut I, $z < 0.4$) and a high-redshift sample (Cut II, $z > 0.4$). Initially, two colour tracks are defined:

$$c_\perp \equiv (r - i) - (g - r)/4 - 0.18, \tag{1}$$

$$d_\perp \equiv (r - i) - (g - r)/8 \approx r - i. \tag{2}$$

Then, the following colour-cuts are applied:

$$\text{Cut I}: \quad |\, c_\perp \,| < 0.2, \tag{3}$$

$$\text{Cut II}: \quad d_\perp > 0.55, \tag{4}$$

$$g - r > 1.4. \tag{5}$$

[1] http://www.2slaq.info/.

The final cut, $g - r > 1.4$, is effective on isolating the sample from the stellar locus. In addition to these selection criteria, all galaxies with $g - r > 3$ and $r - i > 1.5$ are eliminated. These last constraints are helpful on removing stars with unusual colours, without discarding real galaxies (Padmanabhan et al. 2005). However, it is important to keep in mind that a 5 per cent stellar contamination may still be present, as pointed out by Collister et al. (2007) (see section 4 of their paper), who applied similar selection criteria. None the less, their criteria lead to the selection of more objects than the one adopted here.

These colour-cuts are still not enough to select LRGs from SDSS (see discussion in Eisenstein et al. 2001). Therefore, additional cuts in magnitude are applied. First, a colour track which is approximately parallel to the low-redshift locus is defined for Cut I:

$$c_{||} = 0.7(g - r) + 1.2(r - i - 0.18). \quad (6)$$

Then, the following cuts are implemented:

$$\text{Cut I} : r_{\rm Petro} < 13.6 + c_{||}/0.3, \quad r_{\rm Petro} < 19.7; \quad (7)$$

$$\text{Cut II} : i < 18.3 + 2d_{\perp}, \quad i < 20. \quad (8)$$

$r_{\rm Petro}$ is used for consistency with the original SDSS LRG target selection. Except for the numerical values of the magnitude cuts in equation (7), Cut I is identical to the SDSS LRG Cut I. The numerical values for Cut II are chosen to derive a population consistent with the first cut. At the redshift range sampled by Cut II the 4000-Å break is moving through the r band. As a consequence, the r-band K-corrections are very sensitive to redshift. Thus, using the i band for Cut II leads to a more robust selection.

When applying these criteria to select LRGs from the DR5 of SDSS a total of 578 160 galaxies are selected using Cut I and 896 988 through Cut II. The combined sample, after excluding overlapping galaxies, adds to 1459 536 LRGs. This sample is from now on called the photometric sample. Note that the high-redshift sample has approximately 74 per cent of the MegaZ-LRG catalogue (Collister et al. 2007). This last catalogue was selected from the DR4 of SDSS. If the same criteria adopted by Collister et al. (2007) is applied to DR5 the number of LRGs retrieved is ∼1.4 million. So, the high-redshift sample in the current work (Cut II) actually represents ∼64 per cent of the MegaZ-LRG catalogue. That is due to the different criteria employed here. I allow only objects at $1.4 < g - r \leqslant 3$ and $r - i \leqslant 1.5$, while Collister et al. (2007) uses $0.5 < g - r \leqslant 3$ and $r - i < 2$. Besides that, they select galaxies at $d_{\perp} > 0.5$, while here the adopted cut is $d_{\perp} > 0.55$. When imposing that galaxies should have spectroscopic measured redshifts and applying the same criteria as above, there are 197 956 LRGs in SDSS. Out of these, 186 572 are at low redshifts (Cut I) and 11 384 at high redshifts (Cut II). The small number of galaxies with spectra available at high redshift is due to the fact that 2SLAQ was restricted to a small number of fields located in the equatorial stripe of the SDSS survey area. This set is called the spectroscopic sample. Note that this sample does not include stars, as those were removed according to their spectroscopic identification.

3.1 Photometric redshifts of LRGs

The empirical photometric redshift estimators rely on the existence of a training set of objects with spectroscopic redshifts. This set should be representative, in terms of photometry and redshift, of the target sample which will be used later on. The training set used here is the spectroscopic sample (with 197 956 objects) mentioned above. Actually, this sample is divided into 'training' and 'evaluation' samples. I randomly selected 10 000 objects out of the 197 956 LRGs to be the 'training' set. The remaining 187 956 galaxies are kept to play the role of an 'evaluation' sample. The training set is then used to derive an empirical relation between galaxy colours and redshift, which is then applied to the evaluation set. I found that increasing the training sample to 20 000 objects does not represent a meaningful gain in accuracy. Forcing the 'random' selection to have a fixed percentage at $z > 0.40$ (say 60 per cent) also leads to similar results.

On what follows I discuss which colours are best suited for deriving an empirical relation used for photometric redshift estimates. In Fig. 1 I show the variation with redshift of the apparent magnitude r and of five SDSS colours, namely $(u - g)$, $(g - r)$, $(g - i)$, $(r - i)$, $(i - z)$. The data points represent the 10 000 galaxies randomly selected for the training sample. On each bandpass K-corrections are obtained through the convolution of an SED characteristic of early-type galaxies (taken from Coleman, Wu & Weedman 1980, CWW from now on) with the SDSS filters. The expected colours in different redshifts are the result of adding the colour of an elliptical galaxy at zero redshift and the difference in K-corrections between two bands. The zero-redshift colours are taken from table 3 of (Fukugita, Shimasaku & Ichikawa 1995). These colour tracks are shown by the solid lines (red in the electronic edition) of panels (b)–(f) of Fig. 1. A small offset was noticeable between the colour tracks and the data points. I estimated these offsets (a factor of <0.15) and took them in account for the figure. In panel (d) the dotted line (green in the electronic edition) indicates a second-order polynomial fit to the relation between $(g - i)$ colour and redshift. In this panel the dashed line (blue in the electronic figure) shows the result of a fourth-degree polynomial.

A few features are readily noticed from the inspection of this figure. First, $(u - g)$ shows a large scatter and does not follow the colour track expected for an elliptical galaxy (due to the lower sensitivity of the u filter, especially for LRGs). Secondly, the $(i - z)$ colour shows little variation with redshift. Thus, we do not expect these two colours to contribute in a meaningful way for a photometric redshift estimator. Thirdly, the $(g - r)$ colour shows the expected large variation at low redshifts ($z < 0.35$), becoming nearly flat afterwards. The $(r - i)$ colour shows the opposite trend, being nearly constant at $z < 0.35$ and increasing fast for higher z. The $(g - i)$ colour combines the results of the two previous colours, showing a large variation at low z and a less steep dependence at higher z. These results are mainly associated with the shifting of the 4000-Å break between the g and r filters at $z \sim 0.35$. Lastly, it is worth noting that the $(r - i)$ colour has a much smaller scatter in comparison to the ones based on the g band $(u - g, g - r, g - i)$. That happens because the reddest filters are better suited for sampling these types of galaxies.

I then used the training data to estimate different empirical relations between colours and redshifts. These relations are based on a variety of combinations of colours, with the use of the r-band magnitude in a few cases. These relations are then applied to the 187 956 galaxies of the evaluation sample. The redshift accuracy is characterized by the residual between the spectroscopic and photometric redshifts ($\delta_z = z_{\rm spec} - z_{\rm phot}$) and the standard deviation

$$\sigma = \sqrt{\frac{1}{N-1}\sum(\delta_{zi} - \mu)^2}\,, \quad (9)$$

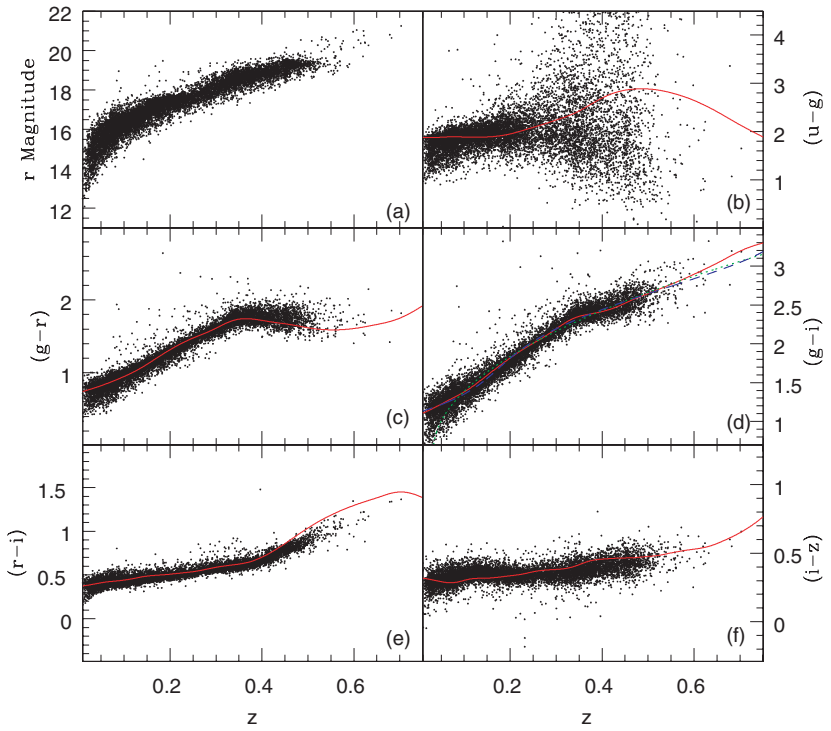

Figure 1. The variation of apparent magnitude r and five SDSS colours with redshift. The six panels show the following parameters versus redshift: (a) magnitude; (b) $(u-g)$; (c) $(g-r)$; (d) $(g-i)$; (e) $(r-i)$; (f) $(i-z)$. The solid line (red in the electronic version) on each panel indicates the expected colour variation of early-type galaxies. In panel (d) the dotted line (green in the electronic edition) indicates a second-order polynomial fit to the relation between $(g-i)$ colour and redshift. In the same panel the dashed line (blue in the electronic figure) shows the result of a fourth-degree polynomial.

where δ_{zi} is the residual for the ith galaxy and $\mu = \langle \delta_z \rangle$ is the mean residual. The sum is performed over all N data points. I also computed the mean and standard deviation (now called μ_0 and σ_0) for the case where the residual is weighted by the factor $1 + z_{\rm spec}$ [$\delta_{z0} = (z_{\rm spec} - z_{\rm phot})/(1 + z_{\rm spec})$]. When computing μ and σ (or μ_0 and σ_0) I only use galaxies with $|\delta_z|$ or $|\delta_{z0}| < 0.10$. This gives equivalent results to reject outliers at the 3σ level. Including the gross outliers increases the standard deviation, as σ is very sensitive to the presence of outliers. For instance, in the fifth row of Table 1 (results based in the colours $g-r, g-i$ and $r-i$) the values of σ and σ_0 are raised from 0.027 and 0.023 to 0.032 and 0.026, respectively.

The mean and standard deviation, as well the fraction of galaxies with a valid $z_{\rm phot}$ (the number of gross outliers is 100 minus this fraction), obtained for different empirical relations are summarized in Table 1. The last row of this table has the results obtained when considering the colour track based on the elliptical template from CWW for the $(g-i)$ colour. In other words, I simply use the track exhibited as a solid line in panel (d) of Fig. 1 to compute redshifts from the observed $(g-i)$ colour. The remaining rows (1–7) in Table 1 show the results for the following polynomial fits:

$$z_{\rm phot} = A + B(g-i) + C(g-i)^2 + D(g-i)^3 + E(g-i)^4,$$
$$z_{\rm phot} = A + B(g-i) + C(r-i),$$
$$z_{\rm phot} = A + Br + C(g-r) + D(r-i),$$
$$z_{\rm phot} = A + Br + C(g-i) + D(r-i),$$
$$z_{\rm phot} = A + B(g-r) + C(g-i) + D(r-i),$$
$$z_{\rm phot} = A + B(g-r) + C(g-i) + D(r-i) + E(r-z),$$
$$z_{\rm phot} = A + B(g-r) + C(g-i) + D(r-i) + E(i-z) + F(g-r)^2 + G(g-i)^2 + H(r-i)^2 + I(i-z)^2.$$

Table 1. The mean, standard deviation and fraction of LRGs with a valid photometric redshift ($z_{\rm phot} > 0$ and $|\delta_{z0}| < 0.10$). All rows show the results of polynomial fits obtained using different parameters. The first row lists the results when applying a polynomial of fourth order to the relation between the colour $(g-i)$ and redshift. In the second row I list the results when using the colours $(g-i)$ and $(r-i)$. Those based on the r magnitude and colours $(g-r)$ and $(r-i)$ are shown in the third row. In the fourth row the results represent the use of the r magnitude and colours $(g-i)$ and $(r-i)$. The fifth row shows the results when using the colours $(g-r)$, $(g-i)$ and $(r-i)$. In the sixth row I show the results obtained when adding the colour $(r-z)$ to the previous set. The seventh shows the result of a second-order polynomial to the colours $(g-r)$, $(g-i)$, $(r-i)$ and $(i-z)$. Details about the fits are given in the text. In the last row I show the results achieved when using the colour track based on the elliptical template from CWW for the $(g-i)$ colour.

Relation	μ	σ	μ_0	σ_0	Fraction (per cent)
Photometric redshift errors of LRGs from different relations					
*gi*4	−0.0024	0.029	−0.0026	0.025	94.1
giri	−0.0008	0.028	−0.0012	0.024	97.1
rgrri	−0.0007	0.027	−0.0008	0.023	96.4
rgiri	−0.0007	0.027	−0.0008	0.023	96.5
grgiri	−0.0075	0.027	−0.0068	0.023	97.4
grgiririz	−0.0068	0.027	−0.0062	0.024	97.1
*grgiriiz*2	−0.0062	0.028	−0.0058	0.025	96.9
TEMP-CWW	0.0016	0.030	0.0013	0.026	99.1

From the inspection of Table 1 we see no large differences among the different relations. It is interesting to note that the results based on the $(g-i)$ colour track (using the SED from CWW) show the highest fraction of galaxies with a valid photometric redshift. However, the

scatter determined in this way is a little larger when compared to other solutions. Besides that fact, the inspection of the $z_{\rm phot} - z_{\rm spec}$ relation shows a poor correlation for $z_{\rm spec} > 0.30$ (for the TEMPCWW results). We also note that the results based on the r-band magnitude and colours have more outliers than the solutions based on two or more colours (but no magnitude). The result based on the r magnitude and colours $(g - r)$ and $(r - i)$ is at the same level of the one that uses $(g - i)$, instead of $(g - r)$. So, for LRGs, the $(g - i)$ does not seem to be superior to the $(g - r)$ for redshift estimation. That is not true for galaxy clusters though (Section 4). I have also tried second- and fourth-order polynomial fits to the $(g - i)$ colour only. However, the overall results were not better. In Table 1 we can see that the fraction of outliers increases for this $(g - i)$ fourth-degree solution. Finally, it is interesting to see that the use of the z filter through the $(r - z)$ or $(i - z)$ colours does not help improving the results. The scatter is at the same level when adding the $(r - z)$ colour, increasing a little if the $(i - z)$ colour is employed with a second-order fit.

I also estimated the uncertainty in the photometric redshift estimates, of each galaxy, through propagation of errors. In this process I consider the error in the coefficients of the empirical fits, as well as the photometric errors for each magnitude or colour. These uncertainties are estimated as shown in the equation below, where I use the fractional uncertainties in the coefficients (such as $A, B, \ldots$) and in the colour and magnitudes used on each fit $(r, g - r, g - i, \ldots)$

$$\Delta z_{\rm phot} = z_{\rm phot} \sqrt{\left(\frac{\Delta_A}{A}\right)^2 + \cdots + \left[\frac{\Delta(g-r)}{(g-r)}\right]^2 + \cdots}. \qquad (10)$$

After inspecting Table 1 and plots of the $z_{\rm phot} - z_{\rm spec}$ relation I decided to consider as the final catalogue of photometric redshifts the one based on the $(g - r)$, $(g - i)$ and $(r - i)$ colours. The coefficients for the empirical relation derived for this case are:

$$A = -0.3068 \pm 0.0006,$$
$$B = \ \ 6.2005 \pm 0.1333,$$
$$C = -5.9933 \pm 0.1331,$$
$$D = \ \ 6.4932 \pm 0.1324.$$

This decision was mainly motivated by the fact that this relation is the one to produce fewer outliers and to show small individual galaxy redshift errors. The results based on relations that involve the magnitude r have large galaxy error estimates, due to the large fractional uncertainties in the magnitude coefficient (Δ_B/B). That is the main reason for not adopting one of these relations. A plot with the comparison between $z_{\rm phot}$ and $z_{\rm spec}$ obtained with the $(g - r)$, $(g - i)$ and $(r - i)$ colours is shown in Fig. 2. Here I plot 15 000 randomly selected points from the evaluation sample (comprising 187 956 galaxies). It is important to mention that these results are in good agreement to what has been found by other authors (Padmanabhan et al. 2005; Collister et al. 2007) and the residuals show no systematic trends with $z_{\rm spec}$. When considering only galaxies with $z_{\rm spec} \leqslant 0.55$, $\sigma = 0.027$ and $\sigma_0 = 0.023$, while for galaxies at $0.55 < z_{\rm spec} \leqslant 0.70$, $\sigma = 0.049$ and $\sigma_0 = 0.040$.

Table 2 presents the LRG catalogue derived from the relation based on $(g - r)$, $(g - i)$ and $(r - i)$ colours. The parameters listed are ra, dec, u, g, r, i, z, u_err, g_err, r_err, i_err, z_err, z_phot, err_z_phot and objID (the object ID within SDSS). The magnitudes are the dereddened model magnitudes. When the estimator led to a negative photometric redshift I set z_phot and err_z_phot to -9.99. That happens for 8509 galaxies (only 0.6 per cent of the total catalogue).

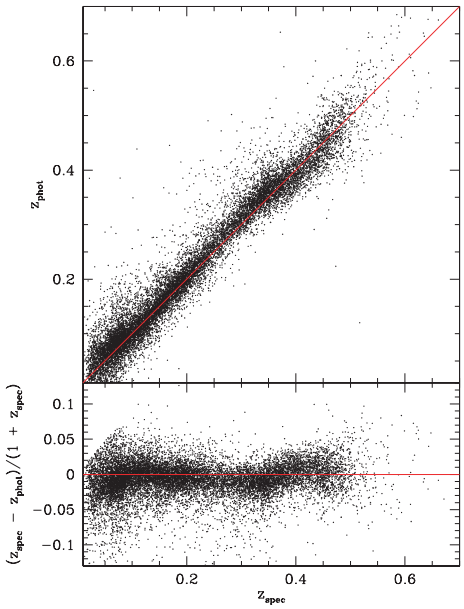

Figure 2. Comparison between $z_{\rm phot}$ and $z_{\rm spec}$ using the empirical relation based on the $(g - r)$, $(g - i)$ and $(r - i)$ colours (upper panel). In the lower panel the residuals $\delta_{z0} = (z_{\rm spec} - z_{\rm phot})/(1 + z_{\rm spec})$ are shown. The solid lines (red in the electronic edition) indicate the $Y = X$ result on top and the zero residual in the bottom panel.

4 GALAXY CLUSTERS

The knowledge of clusters redshifts is essential for estimating other physical parameters of these systems (such as luminosity and richness). Accurate redshifts are also crucial for large-scale structure studies. Perhaps the oldest and simplest way to get photometric redshifts is through the use of single-band galaxy magnitudes within the cluster region. However, the use of colours turned the photometric redshifts of clusters much more accurate in the last years. Colour-based techniques explore the fact that clusters have a large population of early-type galaxies which are characterized by a strong break at 4000 Å. The observation of this feature through different filters shows a well-defined correlation with redshift.

Some cluster selection methods explore this characteristic, estimating a cluster redshift at the time of detection (Gladders & Yee 2005; Koester et al. 2007), while others do the estimation independently of the selection procedure (Gal et al. 2003). Typical accuracy of these colour-based methods is <10 per cent.

The catalogues from Goto et al. (2002) and Koester et al. (2007) are based on SDSS. So, their results provide an optimal basis for comparison to what I find here. Two of the goals aimed in this work are: the derivation of redshifts independent of the selection procedure; and the extension of these estimates to $z \sim 0.55$, when using SDSS data to $r = 21$. The first point is motivated by the fact that some methods for cluster selection [such as the matched filter (MF) or the maxBCG technique] provide a redshift estimate as the output of the selection procedure. Although that is a great advantage, it also makes it difficult to use these methods for redshift estimation of clusters that do not properly *fit* the required properties imposed by these techniques. The photo-z technique presented here is applicable to clusters detected by different methods. I also noted that none of the SDSS cluster catalogues available in the literature have redshift estimates to $z \sim 0.55$. The catalogue of Koester et al. (2007) focuses the regime at $0.1 < z < 0.3$ (some other catalogues

Table 2. Example of the LRG catalogue containing 1459 536 objects. The magnitudes listed are the dereddened model magnitudes. The full table is available online or upon request to the author.

ra	dec	u	g	r	i	z	u-err	g-err	r-err	i-err	z-err	z_phot	err-zp	objID
Example of the LRG catalogue														
0.001 571	14.982 669	26.399	22.074	20.319	19.488	19.079	0.667	0.159	0.050	0.039	0.126	0.472 47	0.067 14	587730774425600743
0.002 045	−9.793 661	22.227	21.282	19.469	18.829	18.490	0.609	0.083	0.025	0.023	0.073	0.389 00	0.034 20	587727179523227851

in SDSS also do not go deeper than that). However, the work of Kim et al. (2002) and Goto et al. (2002) use galaxies at $r \leqslant 21$ and $r \leqslant 21.5$, respectively; and these catalogues are not intentionally driven to a low-redshift regime. The redshift estimates from Kim et al. (2002), based on an MF technique were intentionally truncated at $z = 0.5$ (as they ran the MF up to this redshift). The results from Goto et al. (2002) use the $g - r$ colour track for redshift estimation, but as seen from their fig. 14, their photo-zs are truncated at $z \sim 0.44$, while the spectroscopic sample used for comparison goes to $z = 0.5$ and they probably detect higher redshift systems (considering the magnitude limit adopted). In other words, although accurate they underestimate the redshifts (at least in the high-z regime).

4.1 Selection of a training sample of galaxy clusters

For the determination of photometric redshifts of galaxy clusters the first step that should be taken is the compilation of a list of objects with measured spectroscopic redshifts. Unfortunately, there are not so many clusters with spectra taken at $z > 0.3$, which biases our sample to low-z clusters. Our calibration sample consists of 512 clusters over the area covered by SDSS (DR5). These come from Struble & Rood (1999), Holden et al. (1999), Vikhlinin et al. (1998), Carlberg et al. (1996) and Mullis et al. (2003). The combined sample of these references contains 1805 clusters in the whole sky. After selecting all clusters with redshifts at $0.02 \leqslant z \leqslant 0.55$ and outside $70.0 < \alpha < 110.0$ or $270.0 < \alpha < 300.0$ we are left with a list of 1055 clusters. The right ascension limits are meant to avoid most systems outside the SDSS region. However, many systems that do not overlap with SDSS are still allowed in this list. Then I select data from SDSS for all these clusters. Those falling off the SDSS limits will obviously contain no galaxies. For the remaining I generate finding charts, which are inspected to see if the regions around each cluster (8.0×8.0 Mpc) are well sampled (I exclude clusters with excised regions near their centres). The final list comprises 512 systems. The sample size may not be large enough to properly include evolutionary effects. However, the dominant population of red galaxies in clusters is believed to evolve passively with redshift. So, for clusters, the average colour estimated from these galaxies should provide a clear correlation with redshift. In the future it will be interesting to use larger training samples of clusters for empirical photometric redshift estimators.

The next step is to investigate which colours provide the best connection to spectroscopic redshifts. When doing so I noted that the relations between magnitude or colours to redshift are very well established. However, caution should be taken for the background correction (Section 4.2) and also to the selection of clusters which will be used as the evaluation sample. I noted that most of the 512 clusters can be well represented by the colour tracks of elliptical galaxies. However, there are a few outliers, especially at low redshifts, which should definitely be avoided when training a photo-z estimator. These are mainly associated with wrong redshifts (clusters with few galaxies with available spectra) or projection effects (if there is one or more clusters aligned to a low-redshift system the colour inferred will probably be wrong).

I then decided to gather further information from NASA/IPAC Extragalactic Database (NED) to exclude clusters with a small number of galaxies with redshift available. When doing that I selected from NED all galaxies within 3 arcmin of each cluster centre. After inspecting the information retrieved for each cluster I kept only those with at least three galaxies with a concordant redshift. Besides that, NED also provides a 'special note' for some clusters, meaning that there is some peculiarity with the object (most of times it is a double system or there are different redshifts listed for it). Most of the 512 clusters are from Struble & Rood (1999) who also give the number of galaxies used for measuring the redshift (may be different from above, as the aperture is not 3 arcmin). I then impose that the clusters should not have the 'special note' in NED and have at least three galaxies in Struble & Rood (1999). The final training set comprises 132 systems. Nearly all high-redshift clusters are kept ($z > 0.4$).

4.2 Magnitudes, colours and background correction

In this section I show how I compute mean magnitudes and median colours and apply a background correction along this process. Fig. 3 shows the colour tracks for elliptical (E) galaxies (upper panel) and late-type (Sbc) galaxies (bottom panel) obtained from the convolution of the CWW templates with the SDSS filters. The colours shown are $(u - g)$, $(u - r)$, $(g - r)$, $(r - i)$, $(i - z)$. Other colours are not exhibited to avoid confusion. The analysis shown here is

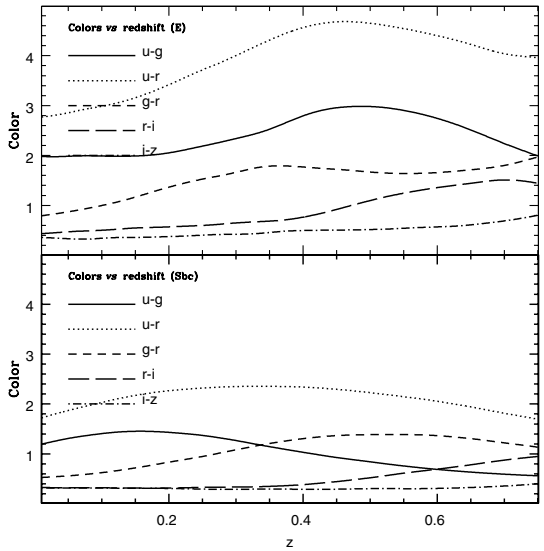

Figure 3. Colour variations as a function of redshift. The top panels show colour tracks for an elliptical SED, while the bottom panel is for a late-type SED (Sbc). Both SEDs are from CWW. The colours exhibited are: $(u - g)$ with solid line; $(u - r)$ with dotted line; $(g - r)$ with short-dashed line; $(r - i)$ with the long-dashed line and $(i - z)$ with the dot–dashed line.

complimentary to the discussion done for the LRGs (Section 3.1). From the top panel it is easy to conclude that the $(i - z)$ colour provides no meaningful information for redshift estimation, while $(r - i)$ colour is expected to be a powerful redshift discriminator at $0.40 < z < 0.60$ and $(g - r)$ works fine at $z < 0.40$. The two colours based on the u magnitude ($u - g$ and $u - r$) show a steep variation with redshift at $z < 0.40$ (with $u - g$ being flat at $z < 0.20$). However, these two colours have a large scatter (Section 3.1) due to the lower signal-to-noise ratio of the u band, specially for early-type galaxies, rendering their use for redshift estimation irrelevant. But that does not prevent us from using the $u - r$ to separate early- and late-type galaxies (see below).

All mentioned so far is in line with the discussion in Section 3.1 (about Fig. 1). Perhaps the most important information that could be extracted from Fig. 3 is a way to discriminate early- and late-type galaxies. We see that the colours $(g - r)$, $(r - i)$, $(i - z)$ are not good for that purpose as the tracks of E and Sbc galaxies are too close and even overlap in some redshift ranges. The other two colours ($u - g$ and $u - r$) seem to be useful for this separation. However, we can note that at low redshifts ($z \sim 0.15$) the difference between E and Sbc galaxies is approximately 0.50 mag only for the $u - g$ colour, while something similar to that does not happen for the $u - r$ colour at a fixed redshift. But it is important to mention that the difference between the lowest value of the $u - r$ colour for ellipticals (at $z = 0$) and the highest for late-types (at $z \sim 0.32$) is of 0.40 mag. So, when using the $u - r$ colour a high-redshift blue galaxy could be mistaken for a low-redshift red galaxy. Even considering that and the already mentioned large scatter of these colours I find that they are helpful for a rough discrimination between galaxy types, which is enough for the purpose of this work. I adopted $(u - r)$ as it shows better results and for being well known for having a bimodal colour distribution (Strateva et al. 2001; Driver et al. 2006).

After this discussion we are now ready to compute mean magnitudes ($r-$band) and median colours ($g - i$ and $r - i$, for instance) for all the 132 clusters of the training sample. For each cluster the number of galaxies as a function of magnitude (N_r) and colours $[N_{(g-i),(r-i)}]$ is determined (within and aperture of 0.50 h^{-1} Mpc, or equivalently 0.71 Mpc for $h = 0.7$). I use bins of 0.10 mag for generating histograms of these counts. Blank fields are used to estimate the background counts. 50 random regions (with $0.^{\circ}5$ radius) are selected in the sky. When computing the background counts I exclude boxes with values outside the boundary determined by the mean $\pm 3\sigma$ of the 50 original boxes. For each magnitude or colour bin the final background value is the mean of counts from the valid background boxes. Then I generate the background magnitude and colour distributions in the same way as done for each cluster, but scaling the counts for the cluster area. All clusters have their magnitude and colour distributions corrected from the background histograms, leading to the net cluster histograms. These are used to compute the mean magnitude and median colours of the clusters.

I tested three different possibilities for obtaining these corrected cluster counts. The procedure is executed exactly as described above, but in the first case I consider all galaxies with the r-band magnitude less than 21 ($m_r \leqslant 21$, which is approximately the star/galaxy separation limit of SDSS) and in the second I count only galaxies with $m_r \leqslant 21$, but with $m^* - 3 \leqslant m_r \leqslant m^* + 3$, where m^* is the apparent characteristic magnitude of clusters. I consider the bright-end values of the double-Schechter cluster luminosity function (LF) obtained by Popesso et al. (2006). They found $\alpha = -1.09$ and $M^* = -20.94$ within R_{200}. This value of M* is converted to the same cosmology used here and to the proper value at $z = 0$ (taking the mean redshift of their sample as $z = 0.1$). Then, for different redshifts I adopted an evolutionary correction to the value of M*, given by $M^*(z) = M^*(0) + Qz$, with $Q = -1.4$ (Yee & López-Cruz 1999). The absolute characteristic magnitude (m^*) is converted to m^* through the application of the distance modulus formula to each cluster redshift (the variation of m^* with redshift is shown in the upper panels of Fig. 4). The third test done is to compare the values obtained when considering all galaxies to the results found when imposing a selection according to the $(u - r)$ colour.

In Fig. 4 I show the dependence of mean apparent magnitude $r_{\rm mean}$ (two upper panels) and two SDSS colours (median values; four lower panels) with redshift. The colours exhibited are $g - i$ and $r - i$. In the top panels the solid line represents the expected variation of the apparent characteristic magnitude of clusters, while in the four lower panels the solid line indicates the expected colour variation of E galaxies. In all panels, every black dot is a cluster (the 132 of the training sample), while the triangles with error bars (red in the electronic edition) indicate the mean values in redshift bins of 0.05. The left-hand panels (a, b and c) show the results for the case where counts are taken at $m^* - 3 \leqslant m_r \leqslant m^* + 3$ and a colour-cut (in $u - r$) is applied for selecting galaxies. The exact values for this cut are chosen after comparing the $u - r$ colour tracks of E and Sbc galaxies in Fig. 3. Initially, I selected only galaxies with $u - r \geqslant 2.40$, which already led to very good results. However, I found that these could still be improved if a variable cut (with redshift) was applied. The choice of this cut affects most of the higher redshift clusters ($z > 0.20$). I finally decided to select only the galaxies with $u - r \geqslant 2.00$ for clusters at $z \leqslant 0.20$, $u - r \geqslant 2.30$ at $0.20 < z \leqslant 0.40$ and $u - r \geqslant 2.45$ at $z > 0.40$. I found these cuts to give the most accurate photometric redshifts. For comparison Strateva et al. (2001) find that early- and late-types can be well separated by a simple colour-cut at $u - r = 2.22$.

In the right-hand panels of Fig. 4 I show the results of not applying one of the constraints mentioned above (the fixed luminosity range $m^* - 3 \leqslant m_r \leqslant m^* + 3$; or the colour-cut in $u - r$). Panel (d) shows the results for $r_{\rm mean}$ when using all galaxies at $r \leqslant 21$ (counts are not restricted to a fixed luminosity range, but the $u - r$ colour-cut is still enforced). Finally, panels (e) and (f) have the median colour variations when I do not impose a colour-cut in $u - r$ (but the fixed luminosity range is still applied).

It is worth mentioning that the further constraints applied above (in luminosity and colour) are intended to improve the background correction, which is done by the subtraction of every cluster histogram (in magnitude and colours) by the background distribution. From panel (a) we can see that the mean cluster magnitudes show a strong variation up to redshift $\sim$0.4. After that the relation tends to become flatter. That is due to the magnitude limit considered for the survey ($r = 21.0$), which renders the cluster counts at $m^* - 3 \leqslant m_r \leqslant m^* + 3$ truncated for high-redshift systems (as $m^* + 3$ extrapolates $r = 21.0$). Applying a correction to the $r_{\rm mean}$ values, to take into account of this truncation, results in no meaningful improvement in the accuracy of the photometric redshifts. When comparing panels (a) and (d) there is a remarkable difference between computing $r_{\rm mean}$ within a fixed luminosity range ($m^* - 3 \leqslant m_r \leqslant m^* + 3$, for instance) or using the full survey limits (I considered all galaxies with $r \leqslant 21$ for panel d). The main effect is the overestimation of the counts at low redshifts ($z < 0.15$), which leads to the flattening of the $r_{\rm mean} - z$ relation in this regime. That happens because when using all galaxies at $r \leqslant 21$ for low-z clusters, we sample magnitudes that are too faint in comparison to the relevant regime of a cluster LF (such as $m^* - 3 \leqslant m_r \leqslant m^* + 3$). Then, the $r_{\rm mean}$ values become biased towards higher values, with also a noticeable increase in the scatter. For clusters at $z > 0.15$ there is no visible difference because $m^* + 3$

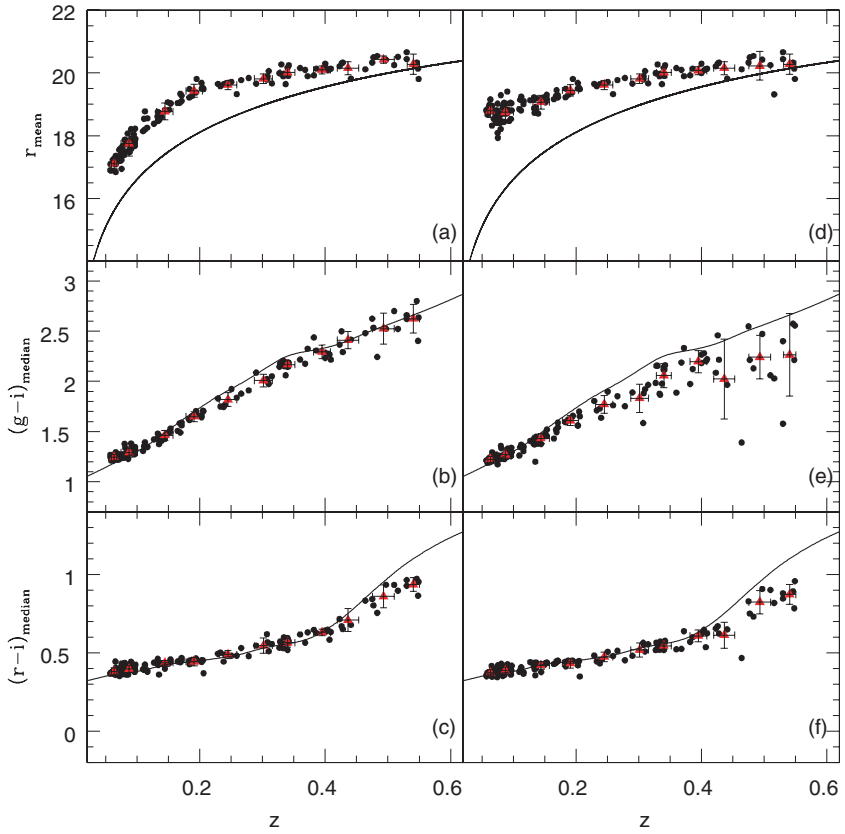

Figure 4. The variation of mean apparent magnitude r_{mean} and two SDSS colours (median values) with redshift. The top two panels show the r_{mean} variation, while the lower four panels exhibit the median colour variations ($g - i$ in the two middle panels and $r - i$ in the lower two). The left-hand panels (a, b and c) show the results when considering counts at $m^* - 3 \leqslant m_r \leqslant m^* + 3$ and with a colour-cut ($u - r$) applied for galaxy selection (this cut is explained in the text). Panel (d) shows the results for the mean magnitude when using all galaxies at $r \leqslant 21$ (counts are not restricted to a fixed luminosity range, but the $u - r$ colour-cut is still applied). Finally, panels (e) and (f) have the median colour variations when I do not impose a colour-cut in $u - r$ (but the fixed luminosity range, $m^* - 3 \leqslant m_r \leqslant m^* + 3$, is still enforced). The solid line on each panel indicates the expected colour variation of early-type galaxies [except for panels (a) and (d) where they indicate the expected variation of the apparent characteristic magnitude of clusters]. Each black dot represents one of the 132 clusters of the training sample, while the triangles (red in the electronic edition) with error bars indicate the mean values in redshift bins of 0.05.

is always close to $r = 21$ (or the survey limit is even extrapolated for high-z systems). A similar discussion, but for richness computation (instead of r_{mean}) is done in Lopes et al. (2006). On what regards colours, the use of all galaxies at $r \leqslant 21$ has no large impact. We only see very few clusters that have their median colours offset from the colour tracks, increasing a little the scatter.

From the comparison of panels (b) and (e) we note a clear trend for underestimation of the $g - i$ colour at $z > 0.15$. The effect is more pronounced at $z > 0.40$. The same effect is noted for the $g - r$ colour (not shown in the plot). For $r - i$ the effect is not too drastic and we only see a mild underestimation at $z > 0.40$. None the less, these results are very useful to show the relevance of imposing a colour-cut (according to $u - r$) for measuring cluster colours. By doing so, we can restrict the analysis to early-type galaxies, rendering the derived colours in good agreement to the expectations of elliptical galaxies.

4.3 Photometric redshifts of galaxy clusters

I then proceed to derive empirical relations to estimate photometric redshifts of clusters. That is done in a similar way to what is shown in Section 3.1 for LRGs, but here I use the mean magnitudes and median colours of clusters, estimated as in the left-hand panels of Fig. 4. In other words, on top of the background corrections I require galaxies to have $m^* - 3 \leqslant m_r \leqslant m^* + 3$ and also impose a colour-cut (in $u - r$), as described above. For the 132 clusters of the training sample I derived the values of r_{mean}, $(g - i)_{median}$ and $(r - i)_{median}$, which are shown in Fig. 4. An empirical relation between these three parameters and redshift is then derived. Other colours are also obtained and will later be used for comparison to the results based on the two above. This relation can be expressed by

$$z_{phot} = A + Br_{mean} + C(g - i)_{median} + D(r - i)_{median},$$

analogous to the ones employed for LRGs. The derived coefficients are:

$$A = -0.4424 \pm 0.0084,$$
$$B = 0.0076 \pm 0.0006,$$
$$C = 0.2382 \pm 0.0024,$$
$$D = 0.2126 \pm 0.0042.$$

To assess the accuracy of the photo-z estimator this relation is applied to all 512 clusters with known spectroscopic redshifts (Section 4.1). However, as we would do with clusters with unknown redshifts, we have to start the procedure with a guessed redshift and iterate it until the photometric redshift difference between two iterations is less than 0.01. A maximum of 10 iterations is allowed.

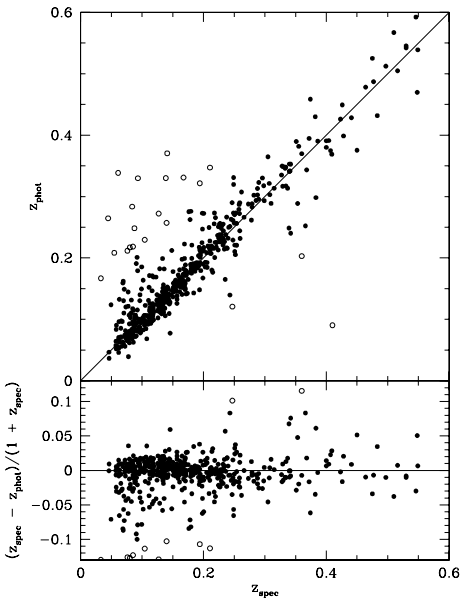

Figure 5. Comparison between $z_{\rm phot}$ and $z_{\rm spec}$ using the empirical relation based on the $r_{\rm mean}$, $(g-i)_{\rm median}$ and $(r-i)_{\rm median}$ colours (upper panel). In the lower panel the residual $[\delta_{z0} = (z_{\rm spec} - z_{\rm phot})/(1 + z_{\rm spec})]$ is shown. The solid lines indicate the $Y = X$ result on top and the zero residual in the bottom panel.

Convergence is not found for only two of the 512 clusters. This iterative procedure is necessary as we sample 0.5 h^{-1} Mpc and $m^* - 3 \leqslant m_r \leqslant m^* + 3$ for each cluster, but we do not know what the redshift of the cluster is (used to determine this radius and luminosity range for galaxy selection). So, we start with a guessed value ($z_{\rm guess} = 0.15$), compute the mean magnitude and median colours and apply the empirical relation obtained above. Then we use the new redshift estimate to repeat the procedure until convergence is achieved. The $z_{\rm phot} - z_{\rm spec}$ comparison is shown in the top panel of Fig. 5, while the weighted residuals are exhibited in the bottom panel. Clusters with $|\delta_{z0}| < 0.10$ (see definition in Section 3) are shown as filled circles (489 or 96 per cent of the 512). The remaining ones are denoted as open circles. If I had used more than 132 clusters (e.g. all the 512) to derive the photometric redshift empirical relation the final results would have a slightly increased scatter. That is due to the fact that some clusters have spectroscopic redshifts derived from a small number of galaxies. That is why it is important to use a clean sample to train the calibration (Section 4.1).

In Table 3 I summarize the results obtained for relations based on different combinations of mean r magnitude and median colours. Analogously to what was done for LRGs, I only use clusters with $|\delta_z|$ or $|\delta_{z0}| < 0.10$ when computing μ and σ (or μ_0 and σ_0). Had I included the gross outliers, the standard deviation for the *rgiri* relation (third row of Table 3) is raised from 0.026 and 0.023 to 0.045 and 0.039 (values of σ and σ_0, respectively). Note that the fraction of outliers is only ∼5 per cent. See also the next section for some considerations regarding accuracy.

The results in this table are first shown for all clusters and then for those at $z \leqslant 0.30$ and $z > 0.30$. In each case, the first four rows list the results considering galaxies at $m^* - 3 \leqslant m_r \leqslant m^* + 3$ and with the $u - r$ colour-cut applied. The fifth row exhibits the values when a fixed luminosity range is not imposed, while the last row has the results without a colour-cut. When considering all clusters (but only the first four rows), we see no large differences, except for the

Table 3. The mean, standard deviation and fraction of clusters with a valid photometric redshift ($z_{\rm phot} > 0$ and $|\delta_{z0}| < 0.10$). All rows show the results of polynomial fits obtained using different parameters. In the first row I list the results when using the colours $(g-i)$ and $(r-i)$. Those based on the r magnitude and colours $(g-r)$ and $(r-i)$ are shown in the second row. In the third row the results represent the use of the r magnitude and colours $(g-i)$ and $(r-i)$. The fourth row shows the results when using the colours $(g-r)$, $(g-i)$ and $(r-i)$. Fifth row exhibits the results obtained with colours $r_{\rm mean}$, $(g-i)_{\rm median}$ and $(r-i)_{\rm median}$, but not being restricted to galaxies at $m^* - 3 \leqslant m_r \leqslant m^* + 3$. All galaxies at $r \leqslant 21$ are used in this case. In the sixth row the results are again for r magnitude and colours $(g-i)$ and $(r-i)$, but the galaxies selected have no $u-r$ cut applied. In the continuation of the table, the same type of information is also given for low-redshift ($z \leqslant 0.30$) and high-redshift clusters ($z > 0.30$).

Relation	μ	σ	μ_0	σ_0	Fraction (per cent)
Photometric redshift errors of clusters from different relations					
giri	−0.0021	0.026	−0.0015	0.023	95.1
rgrri	−0.0056	0.028	−0.0052	0.025	95.3
rgiri	−0.0046	0.026	−0.0040	0.023	94.9
grgiri	−0.0023	0.026	−0.0021	0.023	95.1
rgiri ($r \leqslant 21$)	−0.0021	0.028	−0.0021	0.025	95.1
rgiri (no $u-r$ cut)	−0.0021	0.029	−0.0018	0.026	93.9
Clusters with $z \leqslant 0.30$					
giri	−0.0020	0.024	−0.0019	0.022	95.3
rgrri	−0.0069	0.025	−0.0065	0.024	95.3
rgiri	−0.0050	0.024	−0.0048	0.022	94.9
grgiri	−0.0023	0.025	−0.0025	0.023	95.3
rgiri ($r \leqslant 21$)	−0.0025	0.026	−0.0029	0.024	95.4
rgiri (no $u-r$ cut)	−0.0024	0.026	−0.0024	0.024	96.1
Clusters with $z > 0.30$					
giri	−0.0023	0.037	0.0016	0.030	92.9
rgrri	0.0051	0.041	0.0052	0.031	94.6
rgiri	−0.0007	0.037	0.0026	0.030	94.6
grgiri	−0.0022	0.036	0.0016	0.030	92.9
rgiri ($r \leqslant 21$)	0.0013	0.042	0.0042	0.033	92.9
rgiri (no $u-r$ cut)	0.0010	0.048	0.0045	0.041	76.8

values achieved for the *rgrri* relation ($r_{\rm mean}$, $(g-r)_{\rm median}$ and $(r-i)_{\rm median}$) which have σ (or σ_0) a little higher than the rest. The same is also true for low-redshift clusters (which dominate the sample) and the high-redshift ones. Actually, the fraction of clusters at high z with a valid redshift decreases a little for the *giri* and *grgiri* cases. Considering that, I decided to adopt the redshifts obtained with $r_{\rm mean}$, $(g-i)_{\rm median}$ and $(r-i)_{\rm median}$, as they have fewer outliers and low dispersions at all redshifts. The coefficients for equation (11) are listed above. These results show the importance of using $r_{\rm mean}$ and $(g-i)_{\rm median}$ for photometric redshift estimation at $z < 0.55$ in SDSS. The use of $(g-r)_{\rm median}$ leads to an increased scatter (mainly at high z), while using only colours (without $r_{\rm mean}$) gives lower completeness at high z.

On what regards the results without the luminosity or colour constraints (rows 5 and 6), they are noticeably worst than when these constraints are enforced. When the luminosity restrictions are not applied the results are at a similar level compared to what is obtained in the *rgrri* case (except for the number of outliers at high z). Similar values are also obtained for the case where no colour-cut is applied, but only when low-z clusters are considered. At $z > 0.30$ the standard deviation and fraction of outliers achieved without the

colour-cut are the highest among all. That tells us how important is the selection of early-type galaxies for cluster photo-z estimation at higher redshifts. Without this pre-selection the cluster regions will likely be contaminated by lower redshift sources or, more importantly, by blue galaxies (which have an increased fraction at higher redshifts), thereby increasing the final error and biasing the results to low values. In other words, a simple background correction is not enough, even considering that we are working with a small aperture ($0.50\,h^{-1}$ Mpc).

4.4 Considerations about photometric redshift accuracy

It is important to mention that the redshift accuracy achieved in this work is lower than what other authors found with SDSS data. I estimated $\sigma = 0.024$ for $z \leqslant 0.30$, while Goto et al. (2002) had estimated uncertainties of $\sigma = 0.015$ and Koester et al. (2007) of $\sigma = 0.01$ at $0.10 < z < 0.30$. Note that Goto et al. (2002) show residuals for clusters at $z > 0.08$. The inclusion of lower redshift clusters in our sample ($z > 0.05$) helps increasing the scatter, but by no means can explain the difference to other results.

Koester et al. (2007) estimate redshifts as part of the selection of clusters. Each cluster has assigned to it the redshift of a galaxy (brightest cluster galaxy, BCG) that maximizes the likelihood of representing a cluster centre. Goto et al. (2002) have the redshift estimates done after cluster detection (with the *cut and enhance* method). Their estimates are based on an early version of the maxBCG technique, and are not identical to the ones from Koester et al. (2007). For a given redshift, they start counting the number of galaxies within the cluster detection radius brighter than $M_r = -20.25$ and within ± 1 mag in $g - r$ around the colour prediction for elliptical galaxies (Fukugita et al. 1995). The procedure is repeated for several redshifts in steps of $\delta z = 0.01$. After the background is taken into account on each bin, the redshift of the bin with the largest number of galaxies is considered to be the cluster-estimated redshift.

In this work redshifts are obtained through the application of an empirical relation to $r_{\rm mean}$, $(g-i)_{\rm median}$ and $(r-i)_{\rm median}$. The process is started with a guessed redshift ($z_{\rm guess} = 0.15$) and is iterated until convergence is achieved. For the 512 clusters used here the photometric redshift accuracy is simply given by the comparison of the measured spectroscopic redshift and the photometric estimate. So, it is clear that this process is not guided whatsoever. For comparison, Koester et al. (2007) estimate the accuracy of their estimates by selecting all clusters from their catalogue that have spectroscopic redshifts for their BCGs. For those, they compare $z_{\rm phot}$ and $z_{\rm spec}$, where $z_{\rm phot}$ is the cluster photo-z estimated from the maxBCG algorithm and $z_{\rm spec}$ is the spectroscopic redshift measured for the BCG galaxy in question. That seems a fair comparison, but the authors also recognize that $\sim$16 per cent of their clusters suffer from projection effects, which could affect cluster redshift estimates based on several galaxies (and not only the BCG). On the other hand, Goto et al. (2002) seem to guide the comparison of $z_{\rm phot}$ and $z_{\rm spec}$. In their words: 'the redshift of the SDSS spectroscopic galaxy within the detected radius and with nearest spectroscopic redshift to the estimation is adopted as the real redshift'. Such procedure obviously biases the comparison to small-redshift offsets, based on a single galaxy.

I decide to estimate the residuals in a similar way to what was done by Koester et al. (2007) and Goto et al. (2002). For that purpose I did not consider the 'main' flux-limited sample of SDSS. Instead, I used only the 197 956 LRGs, with spectra available, selected in Section 3. In the first case I selected the nearest galaxy to the cluster centre (within a maximum aperture of 60 arcsec). If the LRG selected is close to the centre it might be the BCG of the cluster. However, it is important to note that I do not make any magnitude or colour requirement for that selection. This simple approach results in $\sigma = 0.021$ (or $\sigma_0 = 0.017$) for the full redshift range of the clusters used here. When I restrict the sample to clusters at $0.10 \leqslant z \leqslant 0.30$ I find $\sigma = 0.014$ (or $\sigma_0 = 0.011$). The last results are closer to Koester et al. (2007) for the same redshift interval. For the full sample there are 111 ($\sim$23 per cent) clusters (out of 489 with a valid photo-z), while there are 71 at $0.10 \leqslant z \leqslant 0.30$.

To perform a comparison to Goto et al. (2002) I did something similar to what they did. The only difference is to use a radius of $0.50\,h^{-1}$ Mpc, instead of the 'detection' radius available within their catalogue. That should not result in meaningful differences as you do not want to select a galaxy that is too far from the cluster centre. So, within $0.50\,h^{-1}$ Mpc I select the LRG with the closest spectroscopic redshift to the value of $z_{\rm phot}$ for each cluster. For 301 clusters ($\sim$62 per cent) there is at least one LRG inside $0.50\,h^{-1}$ Mpc. Out of those, 206 are at $0.10 \leqslant z \leqslant 0.30$. I found $\sigma = 0.018$ (or $\sigma_0 = 0.017$) for the full sample and $\sigma = 0.017$ (or $\sigma_0 = 0.015$) at $0.10 \leqslant z \leqslant 0.30$.

One problem with these tests is the fact that correlating the value of $z_{\rm phot}$ for a cluster with $z_{\rm spec}$ for a single galaxy may lead to wrong matches due to projection effects. As these clusters represent a combination of objects from different catalogues in the literature they are not supposed to have their centroid perfectly matched with a BCG (substructure can affect the centroid determination). That is not the case of the catalogue from Koester et al. (2007). So, if their BCG has a spectroscopic observation and their code works properly, they will have a good correlation between $z_{\rm phot}$ and $z_{\rm spec}$ (except, perhaps, for clusters with strong projection effects). None the less, it is encouraging that the two tests above result in an improved accuracy.

To minimize the influence of projection effects I also estimated the accuracy in a third way. For each cluster I select all LRGs within 360 arcsec from the cluster centre. From those, I check whether the galaxy closest to the centroid is at a maximum distance of 120 arcsec. If it is so, then I assume the redshift of this LRG as a reference ($z_{\rm ref}$). Then, from all the other galaxies (LRGs) selected within 360 arcsec I take those that have a maximum redshift difference $|z_{\rm lrg} - z_{\rm ref}|$ of 0.030. If I end up with at least three galaxies I take the mean of these redshifts to be the value of $z_{\rm spec}$. When comparing these to $z_{\rm phot}$ I find $\sigma = 0.016$ (or $\sigma_0 = 0.018$) for the full redshift range and $\sigma = 0.011$ (or $\sigma_0 = 0.010$) at $0.10 \leqslant z \leqslant 0.30$. There are 125 clusters ($\sim$26 per cent) in the full sample and 81 in the restricted redshift interval. Note that this process is not guided whatsoever. I use an aperture that does not scale with redshift (360 arcsec) and check if there are at least three LRGs at the same redshift, taking as reference the redshift of the LRG closest to the cluster centre (within 120 arcsec). As I impose a minimum number of three galaxies, projection effects are minimized and we can see that the accuracy is greatly improved, reaching the 0.01 level found by Koester et al. (2007).

4.5 Application of the photometric redshift estimator

As an application of the empirical photometric redshift relation obtained in Section 4.3 I used SDSS data to derive new redshift estimates for the supplemental version of the Northern Sky Optical Cluster Survey (NoSOCS; Lopes et al. 2004). This cluster catalogue contains candidates to $z \sim 0.50$, but the redshift estimates were based on a simple magnitude–redshift relation. For that project the

magnitude limit adopted was $r = 21$. Thus, due to the large photometric errors, the use of the $g - r$ colour was not possible with data from the Digitized Second Palomar Observatory Sky Survey (DPOSS). For details see Lopes et al. (2004).

This supplemental version of NoSOCS comprises 9956 cluster candidates over 2700 deg^2. I have recently searched the DR5 of SDSS to see which clusters already have SDSS data. After some catalogue cleaning and inspection of all cluster regions from SDSS I found that 7409 NoSOCS clusters are well imaged. I then derived new photometric redshifts for these systems using the relation obtained in this work. Actually, I got new redshifts without recentring and also considering new coordinates, referred to as luminosity-weighted positions (Lopes et al. 2006). As expected, the new redshift distribution has most of clusters at $z < 0.4$ with a tail extending to $z \sim 0.55$. Details about this updated version of NoSOCS will be given in future work, where substructure and superposition effects will be investigated. For the lower redshift clusters ($z < 0.1$) the updated version will also provide measures of velocity dispersion, virial radius and mass; and this cluster subset will be used for comparison to X-rays, as well to study scaling relations.

4.6 Groups in redshift space with photometric estimates

In the DR5 it is also possible to get photometric redshift estimates of nearly all galaxies in SDSS. Two sources of redshifts are provided. They are simply called photoz and photoz2. The first one is based on template fitting using the technique of Csabai et al. (2003). The second version is provided by the UChicago/Fermilab/NYU group and is based on neural networks, with implementation similar to Collister & Lahav (2004). For the results shown below I used photoz (Csabai et al. 2003).

For the 512 clusters from the training sample I selected photometric redshifts for all galaxies from DR5. The same information was also derived for a region of ∼400 deg^2. The boundaries of this area are $168.4 < \alpha < 191.6$ and $20.0 < \delta < 40.0$, and were chosen to avoid regions close to very nearby clusters, such as Virgo and Coma. This sample will work as a large 'sky' photometric redshift survey.

I investigated the possibility to identify groups in redshift space by simply using these galaxy photometric redshifts. I proceed as follows. First, for each cluster, I start identifying groups in redshift space. For that purpose, I use the gap-technique described in Katgert et al. (1996) and Olsen et al. (2005), which identifies gaps in the redshift distribution that are larger than a given value to separate groups. The gap size adopted is $\Delta z = 0.005(1 + z)$ (Olsen et al. 2005), which is approximately 1500 km s^{-1} in the rest frame. I considered the photometric redshifts available in the training sample to select all galaxies within 0.50 h^{-1} Mpc of the cluster centre. When performing galaxy selection, the only requirement I make is that the photo-z of the galaxy should be greater than 0.01 (to avoid very nearby structures and failures within the SDSS estimates). After applying the gap-technique, a number of groups in z-space is identified for each cluster.

The next step is to assess the significance of each of these groups. For that I consider the area of 400 deg^2 described above. For each group I draw 1000 sets of galaxies from the 400-deg^2 catalogue. These sets have the same number of galaxies as in the cluster region where the group was identified. The gap-technique is applied exactly as before and then I check the probability of finding groups with at least the same number of galaxies at the redshift of the original group. A field group is considered if its redshift is within ± 0.005 of the group identified in the cluster region. The significance is given by the difference between one and the achieved probability. I only consider groups that are significant at the 99 per cent level.

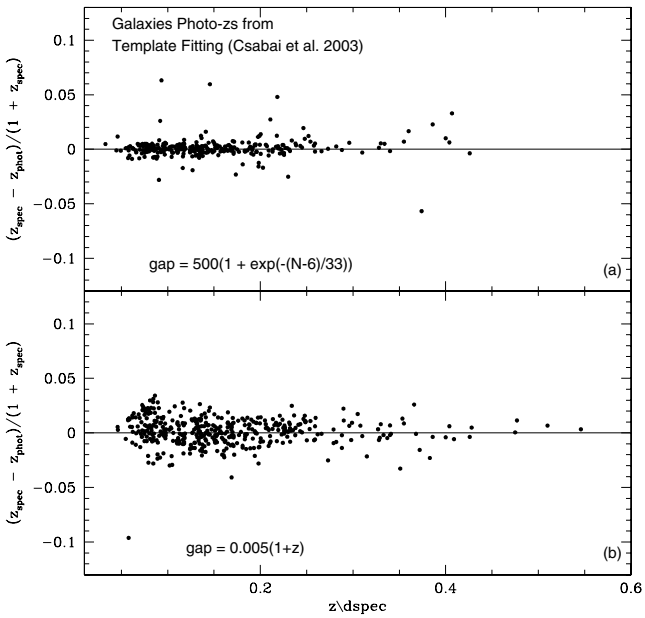

Figure 6. Residual plots between spectroscopic redshifts of training sample clusters and the closest group identified in redshift space using photometric redshifts from SDSS. From the significant groups I select the one with the smallest redshift difference, at least three member galaxies and within 3 arcmin of the cluster centre. The top panel considers the gap size $\Delta z = 500\{1 + \exp[-(N - 6)/33]\}/c$, while for the bottom panel I use $\Delta z = 0.005(1 + z)$. Details are given in the text.

From all the significant groups I select the one that has the smallest redshift difference to the spectroscopic value of the cluster in question. This group should also have at least three member galaxies (most have many more) and be found within 3 arcmin of the cluster centre. I have also run this group identification procedure with one slight modification, which is the gap size. That was modified to $\Delta z = 500\{1 + \exp[-(N - 6)/33]\}/c$, where N is the number of galaxies found in the redshift survey of a cluster (Adami et al. 1998), and c is the speed of light in km s^{-1}.

In Fig. 6 I summarize the results obtained for the two gap sizes adopted. In the bottom panel (b) the gap size considered is $\Delta z = 0.005(1 + z)$, being $\Delta z = 500\{1 + \exp[-(N - 6)/33]\}/c$ in the upper panel (a). The total number of clusters identified and the residuals between $z_{\rm spec}$ and $z_{\rm phot}$ are: (a) 298 and $\sigma = 0.011$ (or $\sigma_0 = 0.009$); (b) 414 and $\sigma = 0.014$ (or $\sigma_0 = 0.013$).

It is interesting to see that these galaxy photometric redshifts provide values accurate enough to trace the local environment of clusters. Similar results are achieved when I adopted the photometric redshifts of LRGs (Section 3). From the inspection of Fig. 6 it is clear that the identification of groups at higher redshifts ($z > 0.4$) becomes very hard. That is due to the poor sampling of the LF of clusters at this redshift interval for $r \leqslant 21$. When using the gap size $\Delta z = 0.005(1 + z)$ we identify more clusters than with the other gap, but at the cost of having an increased scatter.

Considering the larger number of blue galaxies at high redshifts the red sequence in some clusters may be ill defined. So, it is important to consider alternative approaches to cluster detection, such as the selection based on galaxy photometric redshifts.

5 CONCLUSIONS

In this work I described the construction of a large catalogue of LRGs with photometric redshifts at $z < 0.70$ within SDSS.[2] This catalogue is based on an empirical relation to derive the photo-z values. Such relation uses three bands only for achieving results as accurate as those obtained by other methods (sometimes based on more bands; Padmanabhan et al. 2005; Collister et al. 2007). The comparison of photometric and spectroscopic redshifts shows no large systematic effects in the redshift range probed, which suggests that this sample is suitable for large-scale structure studies.

I have also investigated what the main systematics are in the estimation of photometric redshifts of galaxy clusters at $z < 0.55$. That represents an improvement respective to some deep cluster catalogues derived from SDSS. When these catalogues have accurate photometric redshift estimates, these are truncated at $z < 0.44$ (Goto et al. 2002). When they go a little further ($z = 0.5$) the estimates are not as accurate (Kim et al. 2002). I also showed the relevance of using the $g - i$ colour, and to a lesser extent the mean r magnitude, to improve the photometric redshift accuracy, especially at high z ($z > 0.4$). Besides that, I show that on top of a 'traditional' background correction it is very important to select galaxies from a fixed luminosity range and perform a careful selection of early-type galaxies. In this work, this pre-selection of red galaxies is made through a variable cut (in redshift) in the $u - r$ colour. The main advantage of this type of selection is to only require that clusters should exhibit a population of early-type galaxies towards their cores. There is no need for these galaxies to exhibit a narrow red sequence. This type of requirement is very important for the cases where the photometric errors are large or the red sequence is still being formed, which could be the case at high redshifts.

The results obtained for clusters are independent of the way these are selected. So, the methodology described in this work should be valid for any type of clusters, selected by different techniques and wavelengths. The only requirement is to have the proper filters for separating early- and late-type galaxies and to track the 4000-Å break. For SDSS data this method works for clusters at $z < 0.55$. Other considerations are made regarding accuracy. I show that the results shown here are in good agreement to previous works.

The empirical relation derived for clusters is applied to 7409 clusters from the NoSOCS supplemental catalogue (Lopes et al. 2004) which are found within SDSS. For these clusters I was able to update the photometric redshift estimates, deriving more accurate values than before (when using only magnitudes within the DPOSS data). This cluster catalogue, with the new redshifts, will be updated in a future work, where substructure and superposition effects will be investigated. This catalogue is also being used to derive velocity dispersions and mass estimates for the lower redshift systems ($z < 0.1$) and to investigate scaling relations in clusters.

Finally, I tried to identify groups in redshift space using photometric redshifts of galaxies available in SDSS. I found that for ∼60 per cent of clusters (mostly at $z < 0.4$) it is possible to clearly identify a group using only photometric redshifts of galaxies. When comparing $z_{\rm spec}$ of clusters to the photometric redshift of the nearest group, $\sigma = 0.011$. If a different gap size is employed when searching for the groups the rate of identified systems increases to ∼80 per cent, but the accuracy is a little worse ($\sigma = 0.014$). This procedure represents an alternative approach for cluster detection, based on galaxy photometric redshifts.

[2] The LRG catalogue can be retrieved from the electronic edition of this journal or by request to the author.

ACKNOWLEDGMENTS

PAAL was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processes 03/04110-3 and 06/57027-4). I also thank the referee for a careful reading and comments, which helped the presentation of this work.

This research has made use of the NED which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the participating institutions. The participating institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Adami C., Mazure A., Biviano A., Katgert P., Rhee G., 1998, A&A, 331, 493
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Budavári T., Szalay A. S., Connolly A. J., Csabai I., Dickinson M., 2000, AJ, 120, 1588
Butcher H., Oemler A., Jr, 1984, ApJ, 285, 426
Cannon R. D. et al., 2006, MNRAS, 372, 425
Carlberg R., Yee H., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C., 1996, ApJ, 462, 32
Coleman G. D., Wu C. C., Weedman D. W., 1980, ApJS, 43, 393 (CWW)
Colless M. et al., 2001, MNRAS, 328, 1039
Collister A. A., Lahav O., 2004, PASP, 116, 345
Collister A. A. et al., 2007, MNRAS, 375, 68
Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, AJ, 110, 2655
Csabai I. et al., 2003, ApJ, 125, 580
Driver S. P. et al., 2006, MNRAS, 368, 414
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Fukugita M., Shimasaku K., Ichikawa T., 1995, PASP, 107, 945
Gal R. R., de Carvalho R. R., Lopes P. A. A., Djorgovski S. G., Brunner R. J., Mahabal A. A., Odewahn S. C., 2003, AJ, 125, 2064
Gladders M. D., Yee H. K. C., 2005, ApJS, 157, 1
Glazebrook K. et al., 2007, preprint (astro-ph/0701876)
Goto T. et al., 2002, AJ, 123, 1807
Holden B., Nichol R., Romer A., Metevier A., Postman M., Ulmer M., Lubin L., 1999, AJ, 118, 2002
Katgert P. et al., 1996, A&A, 310, 8
Kim R. et al., 2002, AJ, 123, 20
Koester T. A. et al., 2007, ApJ, 660, 239
Lopes P. A. A., de Carvalho R. R., Gal R. R., Djorgovski S. G., Odewahn S. C., Mahabal A. A., Brunner R. J., 2004, AJ, 128, 1017

Lopes P. A. A., de Carvalho R. R., Capelato H. V., Gal R. R., Djorgovski S. G., Brunner R. J., Odewahn S. C., Mahabal A. A., 2006, ApJ, 648, 209
Mullis C. et al., 2003, ApJ, 594, 154
Olsen L. F., Benoist C., da Costa L., Hansen L., Jørgensen H. E., 2005, A&A, 435, 781
Padmanabhan N. et al., 2005, MNRAS, 359, 237
Popesso P., Biviano A., Böhringer H., Romaniello M., 2006, A&A, 445, 29
Strateva I. et al., 2001, AJ, 122, 1861
Strauss M. A. et al., 2002, AJ, 124, 1810
Struble M. F., Rood H. J., 1999, ApJS, 125, 35
Vikhlinin A., McNamara B., Forman W., Jones C., Quintana H., Hornstrup A., 1998, ApJ, 502, 558
Yee H., López-Cruz O., 1999, AJ, 117, 1985
York D. G. et al., 2000, AJ, 120, 1579

APPENDIX A: SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article.

Table 2. The LRG catalogue. Please note this file is almost 70 MB and may take some time to download.

This material is available as part of the online paper from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2966.2007.12203.x
(this link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TEX/LATEX file prepared by the author.